DOCUMENT RESUME

ED 114 408 TM 004 887

AUTHOR Jongsma, Eugene A.; Warshauer, Elaine
TITLE The Effects of Instruction in Test-Taking Skills upon Student Performance on Standardized Achievement Tests. Final Report.
INSTITUTION New Orleans Univ., La. Dept. of Elementary and Secondary Education.
SPONS AGENCY Louisiana State Univ. Foundation, Baton Rouge.
PUB DATE Jun 75
NOTE 67p.

EDRS PRICE MF-$0.76 HC-$3.32 Plus Postage
DESCRIPTORS *Achievement Tests; Comparative Analysis; Disadvantaged Youth; *Educational Research; Elementary Education; Grade 5; Guessing (Tests); *Instructional Programs; Literature Reviews; *Program Effectiveness; Reading Tests; Research Needs; Standardized Tests; Student Motivation; Test Bias; Test Construction; Testing Problems; Test Reliability; Test Results; Test Validity; *Test Wiseness

ABSTRACT

The purpose of this report is to: (1) summarize the attempts that have been made to define and investigate test wiseness, particularly those efforts made at teaching test wiseness; (2) identify gaps and deficiencies in the existing body of research; (3) make recommendations as to the direction that future research in this area should take; and (4) report the results of two experimental efforts to teach test wiseness. As part of this project, two studies were conducted to determine the effects of a course in test taking skills on the reading achievement scores of two classes of fifth grade inner-city students. The students were divided, based on their reading test scores, into two groups: low achievers and high achievers. These groups were then divided into experimental and control groups. The students were introduced to the program unit which was designed to keep them motivated as well as develop skills in test taking. At the end of the course, the students were administered the Stanford Achievement Test-Reading and the Comprehensive Test of Basic Skills, Level II. The results of the study showed that although students who received instruction in test taking skills average higher scores, their gains were not statistically significant. (Author/DEP)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED114408

FINAL REPORT

THE EFFECTS OF INSTRUCTION IN TEST-TAKING SKILLS UPON STUDENT PERFORMANCE ON STANDARDIZED ACHIEVEMENT TESTS

Eugene A. Jongsma
Project Director

Department of Secondary and Elementary Education
University of New Orleans
New Orleans, Louisiana

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Elaine Warshauer
Research Assistant

The research reported herein was performed pursuant to a grant with the Louisiana State University Foundation. Points of view or opinions stated in the report do not necessarily represent official Louisiana State University Foundation position or policy.

June, 1975

TM004 887

FINAL REPORT

THE EFFECTS OF INSTRUCTION IN TEST-TAKING SKILLS UPON STUDENT PERFORMANCE ON STANDARDIZED READING ACHIEVEMENT TESTS

Eugene A. Jongsma
Project Director
Department of Elementary and Secondary Education
University of New Orleans
New Orleans, Louisiana

Elaine Warshauer
Research Assistant

The research reported herein was performed pursuant to a grant with the Louisiana State University Foundation. Points of view or opinions stated in the report do not necessarily represent official Louisiana State University Foundation position or policy.

June, 1975

TABLE OF CONTENTS

Table of Contents i

Preface ii

List of Tables iii

Introduction 1

Review of the Literature. 3

Report of the Project Studies 23

Study I: Suburban Elementary 24

Study II: Inner-City Elementary 32

Discussion of Overall Results 37

References 41

Related Sources 44

Appendix 45

PREFACE

Conducting educational research in the field is often a trying and perplexing experience. Without the assistance and cooperation of principals, classroom teachers, and students, the best laid plans of researchers would be futile.

I am grateful to Jefferson Parish and Orleans Parish school systems for their willingness to cooperate on this project. Specifically, I am indebted to the principals and helpful teachers of Greenpark Elementary School and Mc Donogh #39.

I would also like to express my appreciation to Elaine Warshauer for her role in administering the tests, analyzing the data, and preparing the final report. The assistance of Jill McGovern with the testing is also appreciated.

E.A.J.

LIST OF TABLES

Table		Page
1	Summary of Testwiseness Literature	19
2	SAT Raw Score Means, Standard Deviations, and Differences Between Means With Breakdown By Treatment and Achievement Levels	27
3	Analysis of Variance of SAT Scores for Treatment Groups by Achievement Levels	28
4	Analysis of Variance of SAT Scores of Experimental Group for Teacher (Unit Administrator) by Achievement Levels	29
5	CTBS Raw Score Means, Standard Deviations, and Differences Between Means and Breakdown by Treatment Levels	34
6	Analysis of Variance of CTBS Scores by Treatment Levels	35

INTRODUCTION

The role of testing in American education is becoming increasingly more important. Some authorities have even compared the emphasis on testing reached during the 1970's with that of the 1920's, the decade that spawned the testing movement in the U. S. Tests are now used for a variety of purposes including selection, diagnosis, assessment of student achievement, assessment of teacher effectiveness, and the evaluation of experimental programs.

With the increased emphasis and use of educational testing has come renewed public pressure and concern (Thorndike, 1971). The culturally different, who have not scored well on standardized tests, feel they are biased; classroom teachers, who may be held accountable have a renewed interest in their composition and use; school boards, whose sources of funds are linked to test results, are vitally concerned; and individuals, aware of the growing number of important decisions in their lives that are dependent upon test results, are skeptical.

Because of its importance, reading, more than any other aspect of the curriculum, has received the most attention and criticism in regards to testing. A recent editorial in the Los Angeles Times states, "Much time, effort, and money have been wasted in California on administering reading tests whose results are not only misleading but damaging" (Chasman, 1972). The critics of testing have based their charges on many different issues. Standardized tests, they claim, (1) may not truly assess what students know or what teachers teach; (2) give an

unfair advantage to students who guess; (3) may reflect testwiseness and motivation; (4) contain items that are culturally biased; and (5) use norms that are not appropriate to certain population groups. The limitations of standardized reading tests have been well documented by Farr (1969). In general, most of the attacks on testing question the validity of the tests.

Virtually all authorities in educational measurement recognize testwiseness as a source of variance on educational tests. Most would also agree that variance due to testwiseness is undesirable in that it reduces the validity of the test. The suggestions that are usually made for minimizing this undesirable variance include (1) constructing tests that are "testwise proof," i.e. have clear instructions and a minimum of secondary cues; and (2) providing training in testwiseness so that some students will not have an unfair advantage.

The purpose of this report is to: (1) summarize the attempts that have been made to define and investigate testwiseness, particularly those efforts made at teaching testwiseness; (2) identify gaps and deficiencies in the existing body of research; (3) make recommendations as to the direction that future research in this area should take; and (4) report the results of two experimental efforts to teach testwiseness.

REVIEW OF THE LITERATURE

General Orientation and Definitions

The definition of testwiseness that is most prevalent in the literature is that offered by Millman, Bishop, and Ebel (1965, p. 707). According to these authors, "testwiseness is defined as a subject's capacity to utilize the characteristics and formats of the test and/or the test-taking situation to receive a high score...and is logically independent of the examinee's knowledge of the subject matter for which the items are supposedly measures."

In an effort to provide a theoretical framework to guide future research in this area, Millman, et al. divide the principles of testwiseness into the following two major categories:

I. Elements which are <u>independent</u> of test constructor or test purpose
   - A. Time - using strategies
   - B. Error - avoidance strategies
   - C. Guessing strategies
   - D. Deductive reasoning strategies

II. Elements which are <u>dependent</u> upon test constructor or test purpose
   - A. Intent consideration strategies
   - B. Cue - using strategies

An explanation of the Millman framework is helpful in understanding the literature on testwiseness. The first category includes those principles which would be potentially valid in all testing situations regardless of previous contact with the test constructor or similar tests.

A. <u>Time using strategies</u> apply to tests having time limits and involve procedures designed to use time most efficiently.

B. <u>Error-avoidance strategies</u> apply to all testing situations. They are concerned with the avoidance of careless mistakes by such techniques as paying attention to directions and checking answers.

C. Guessing strategies involve situations in which an examinee may receive credit for answers given on a chance basis.

D. Deductive reasoning strategies are methods used to obtain the correct answer indirectly or when only part of the knowledge needed is present. It would also include use of related data from other parts of the test.

The second main category of the framework, elements which are dependent upon test constructor or test purpose, includes those strategies which are valuable after gaining knowledge of the views of the test constructor or test purpose, or having had contact or feedback from similar tests.

A. Intent consideration strategies involve attention to the intent of the test constructor in including specific questions or items.

B. Cue-using strategies permit one to profit from various cues that may be present in the test when the answer is not directly known. Such cues may include: (1) idiosyncracies of the test constructor such as including more true than false items in true-false tests; (2) differences in the the length of options in multiple-choice tests; (3) grammatical inconsistencies between options and stem; (4) the relevance of specific details; and (5) resemblances between the options and an aspect of the stem.

Stanley (1971) takes a somewhat different point of view in classifying testwiseness as a source of variance. He contends that testwiseness represents systematic variance, but if unrelated to the

criterion of interest, will reduce the validity of the test. To the extent that testwiseness represents a general, somewhat lasting quality of the individual, he believes that it contributes "true" variance. This variance may be difficult to eliminate, according to Stanley, insofar as it may be closely related to basic intellectual activity. The results of some of the research studies to be reviewed in this report raise serious doubts about the relationship between testwiseness and intelligence, the cornerstone of Stanley's position.

Vernon (1962) presents a table of sources of test variance partly derived from Thorndike. He identifies sophistication, sets arising from the student's understanding of instructions as well as other response sets, and guessing, as factors contributing to testwiseness.

Ebel and Damrin (1960) view testwiseness as a specific cognitive skill that is capable of being developed through experience. They consider it to be one of the four bases from which examinees could respond to objective test measures. The other three are: (1) knowledge of content; (2) response set; and (3) chance guessing. Ebel (1965) expresses concern over excessive test-taking skills which enable some examinees to score well on a test for which they are totally unprepared. However, he considers that more error in measurement is likely to result from students who have too little rather than too much skill in taking tests.

This brief review of expert opinion shows that there is no consensus as to the definition of testwiseness or to the factors which comprise the construct of testwiseness. Millman, Bishop, and Ebel's definition and theoretical framework has been the most widely accepted

and has served as a point of departure for a number of investigations. In general, most experts believe that testwiseness does exist and is probably teachable. It is also thought that the elimination or reduction of differences in test-taking skills will provide truer estimates of the knowledge and abilities being assessed.

This review of the literature has been grouped into six categories. Although the categories are not discrete but may be overlapping, they are intended to help the reader organize and conceptualize the work that has been reported. The treatment of some categories is extremely brief simply because little empirical study has been conducted in some of the areas. The categories are as follows:

1. Studies investigating testwiseness techniques independent of the test constructor.
2. Studies investigating testwiseness techniques dependent upon the test constructor.
3. Studies investigating differential effects of age, race, or other specific factors on testwiseness.
4. Studies investigating the relationship between testwiseness and general intelligence.
5. Studies investigating methods of teaching testwiseness.
6. Studies investigating the effect of testwiseness on test reliability and validity.

Studies Investigating Testwiseness Techniques Independent of the Test Constructor

Callenbach (1971) found that second grade students who received instruction and practice in test-taking techniques twice a week for four weeks, achieved significantly higher scores on standardized reading tests. Significant differences were found both on an immediate posttest administered the week following treatment and a delayed posttest administered four months after treatment.

Oakland (1972) devised curricular materials to improve testwiseness for an experimental group of Head Start students. After working with the materials twice a week for six weeks, the experimental group scored significantly higher than the control group on the Matching subtest and Total score of the Metropolitan Readiness Test administered immediately after training. However, group differences four months later were not significant. These results contrast with Callenbach's results which showed a long term effect. Oakland also analyzed his data for differences between high and low performers and found significance on only one subtest, that of Word Meaning. He therefore concluded that when improvement in testwiseness does occur, the improvement is generally not limited to subjects within a particular aptitude range.

Gaines and Jongsma (1974), in a study of fifth grade subjects, found that a one hour unit on test-taking skills administered the day prior to the administration of a standardized reading achievement test yielded significantly higher scores for the experimental group.

Wahlstrom and Boersma (1968) used intervention techinques drawn from both categories of the Millman framework. Ninth grade Canadian students received four 25-minute lectures and a study period in which to review principles of testwiseness. Two posttests were used. Both consisted of social studies content, however, one was poorly constructed and the other well constructed. Significant differences were found for students taking the poorly constructed posttest but not for students taking the well constructed posttest.

Sause and Grieco (1973) found that two 40-minute instructional sessions were effective in producing significantly higher scores on the Otis-Lennon Mental Ability Test for inner-city sixth graders.

Many studies have been made in the areas of guessing and answer changing. For the most part, these studies are outside the realm of this report, however, it's worth noting a couple of specific examples. Bauer (1971) and Slakter (1969) in the field of guessing and Jacobs (1971) and Lynch (1972) in the area of answer changing all found significant increases in test scores after intervention which advised the use of these specific techniques. The consensus seems to be that students will gain by the use of both of these techinques.

The effect of improving test-taking skill by the use of repeated testing with a variety of tests was investigated by Kreit (1967) and Lewis and Biggs (1971). Kreit limited his investigation to intelligence testing of third graders. He found immediate gains in favor of such treatment but no significant difference five months later. Lewis and Beggs used two forms of the Lorge-Thorndike Intelligence Test with sixth graders. They found verbal gains related to the practice effect and nonverbal gains related to recall of specific items.

In summary, the studies reported in this section yielded favorable results regarding the teaching of test-taking skills. The following conclusions can be drawn from these studies:

1. Testwiseness techniques that are independent of the test constructor can be effectively taught to students of all ages, from preschool through college.

2. A wide variety of treatments have been effective, including written instructional materials, verbal lectures, and repeated testing.

3. Improvement in testwiseness has been demonstrated on a variety of posttests, including group intelligence tests, standardized achievement tests, and project-developed tests.

4. While the immediate benefits of instruction in testwiseness appear convincing, the long-term, lasting effects have yet to be consistently demonstrated.

Studies Investigating Testwiseness Techniques Dependent Upon Test Constructor

Studies reported in this section would subscribe to the definition put forward by Diamond and Evans (1972) and Gibb (1964) that testwiseness is "the ability to respond advantageously to multiple choice items containing extraneous clues and to obtain credit on those items without knowledge of subject matter" (p. 145). Diamond and Evans were not attempting to teach testwiseness but to determine whether or not it was present in a group of sixth grade students and, if so, whether it was related to some general ability or was clue specific. Results indicated that testwiseness was present for students of this age and that many students were able to verbalize its presence. Their data further indicates that testwiseness is not some general trait, but rather is specific to the particular clue or cue under investigation.

Ferrell (1972) developed an instrument to determine the degree of testwiseness possessed by high school students. Using five samples of students from five different high schools, it was found that testwiseness was a factor in four out of five groups tested using teacher made tests, and that testwiseness was related to standardized achievement test scores. Results were mixed in an attempt to establish a relationship between testwiseness and grade point average or sex.

Gibb (1964) concluded, in a study with college students, that there are individual differences in testwiseness which are composed, at least in significant part, of the ability to respond advantageously to the presence of "secondary cues" in multiple choice test items. He also determined that secondary cue response could be improved with training. Gibb further states that there is reason to believe that testwiseness is a characteristic of some generality in an unsophisticated population, that it represents a source of invalid variance, and that there is a significant spread of reliable individual differences in testwiseness.

Two studies were reported by Slakter, Koehler, and Hampton. One (1970b) involved high school seniors. This study employed methods to teach testwiseness and guessing to two groups, each of which acted as a control for the other. Results indicated that the testwiseness group achieved significantly higher scores on the measure of testwiseness developed by the authors and administered after training than did the guessing group, which answered significantly more questions on the instrument measuring willingness to guess than did the testwise group. Slakter, et.al. suggests that students be taught testwiseness to reduce errors of measurement and also to reduce the handicap of unsophistication in test-taking. The second report by Slakter, et.al. (1970a) will be discussed in the following section.

In a study with college students, Moore(1971) found that significantly higher mean scores were obtained by an experimental group over a control group on a test of analogies, after exposure to instruction which familiarized them with question formats and types of analogies.

Woodley (1972) conducted a study with adults preparing to take the examination of the American College of Life Underwriters. A number of strategies from both categories discussed by Millman were taught to one group; a test battery given a second group; and no intervention administered to a third group. After completion of the training, a test-wise battery was administered, followed by the A. C. L. U. examination. Results indicated significant differences, favoring the experimental groups, on the testwise battery but not on the final A. C. L. U. exam.

In summary, the studies reported in this section were primarily conducted to determine if subjects were "wise" to various test-taking skills and if such techniques could be successfully taught. A diverse range of strategies were considered including (1) knowledge of the idiosyncracies and intent of the test constructor; (2) cues which might be used, sometimes referred to as secondary cues; (3) relevance of specific details; (4) use of specific determiners; and (5) resemblances between the options and the stem. Overall results suggest that such aspects of test-wiseness can be assessed and can also be taught.

Studies Investigating the Differential Effects of Age, Race, or Other Specific Factors on Testwiseness

Slakter, et. al. (1970a) developed and administered a measure of testwiseness in order to determine the relationship between testwiseness and grade level and testwiseness and sex in grades five through eleven. Selected aspects of testwiseness, such as cue-using strategies and deductive reasoning strategies, were included in their measure. Significant grade level differences in testwiseness were found. Students in

grades five to seven recognized stem-options and absurd-options, but only at grade eight and above were similar-options and specific determiners used effectively. Neither sex effects nor sex by grade interaction effects were significant.

In another study reported previously, Ferrell (1972) found no conclusive relationship between sex and testwiseness in the high school population he studied.

Two studies dealt specifically with the culturally different. Buchanan (1968) analyzed the popularity of distractors for items answered incorrectly between deprived and non-deprived students. He found that indiscriminate guessing was related to lack of information rather than to differences in motivation. Although in unmatched questions, the deprived group showed significantly more guessing than the non-deprived group, on matched questions (those marked incorrectly by both groups), there was no significant difference between the two groups.

Tinney (1968) analyzed the effect of training in test-taking skills on the reading test scores of fifth grade students of high and low socioeconomic levels (SES). He found no significant differences in performance between those receiving the training and those not receiving it. Although statistical significance was not obtained, there was a strong tendency favoring the experimental group. The low SES experimental group made the greatest proportionate gain, although this was not statistically significant. Tinney believes, however, that this latter tendency toward interaction supports his original hypothesis that low SES children would benefit more than high SES children from training in test-taking skills.

Any conclusions that are drawn from the studies in this category must be viewed as extremely tentative because of the scant evidence regarding correlates of testwiseness. On the basis of the studies reviewed in this section, it appears that:

1. Sex is independent of testwiseness
2. Testwiseness generally increases with grade or age level
3. Training in testwiseness is no more effective for students of any particular socioeconomic level, although some investigators continue to hypothesize that such training will be more beneficial to low socioeconomic level students.

Studies Investigating the Relationship Between Testwiseness and General Intelligence

Dunn and Goldstein (1959) made an effort to explore the relationship between testwiseness and general mental ability by examining correlations between scores on blocks of items written in varying degrees of conformance to item writing principles and scores on Army Aptitude Area I, the Army's measure of over-all mental ability. A total of 832 enlisted Army trainees took special forms of the Basic Military Subjects Test and the Mental Ability test. The correlations were homogeneous across item blocks (good vs. poor items), casting doubt on the assumption that intelligence and testwiseness are related. The authors conclude that the "results suggest that the ability to pick up cues on the type of material tested may be found at all levels of intelligence". (p. 178)

In a study discussed earlier, Kreit (1967) investigated whether testwiseness can be taught to relatively test-naive third grade students by exposing them to the experience of taking several different group intelligence tests. He found no significant relationship between pupil intelligence and the learning of test-taking skills.

Diamond and Evans (1972) utilized a test instrument made up of fictitious material to explore various cognitive correlates of testwiseness with a group of sixth grade students. Aspects of testwiseness, defined as secondary cue response, were correlated with other cognitive measures, namely the Lorge-Thorndike Intelligence Test and the Iowa Test of Basic Skills. Factor analysis and examination of the correlational matrices indicate that testwiseness is not some general trait, but rather is specific to the particular clue or cue under investigation. The results of this study support the findings of Dunn and Goldstein (1959).

The findings of the three studies reported in this section are consistent. Testwiseness does not appear to be a global, general trait but is composed of a network of specific and independent skills. Furthermore, the ability to use test-taking strategies is not related to general intelligence as measured by group intelligence tests.

Studies Investigating Methods of Teaching Testwiseness

Although a variety of methods have been successfully used to teach testwiseness, only one study could be found that directly addressed itself to assessing the effectiveness of various methods of training (Langer, Wark, and Johnson, 1973). In the first phase of this study, the authors constructed an instrument to assess the testwiseness of college students. They found that testwiseness does exist in varying degrees among college students and that it can be improved. The second phase of the study was conducted to determine the best method of instruction in testwiseness. Using a four group design, the following treatments were administered:

(1) a lecture on ten major testwiseness cues; (2) a self-instructional programmed booklet; (3) a paper entitled, "A Script to Teach Testwiseness," which was read independently; and (4) no instruction in testwiseness. When students were assessed on the post-Testwiseness Test, all three instructional groups scored significantly higher than the control group. However, there were no significant differences among the three methods of instruction that were employed. It appears that instruction in testwiseness can be achieved by reading, lecture, or programmed exercises with similar results.

Studies Investigating the Effect of Testwiseness on Test Reliability and Validity

In a study discussed previously, Dunn and Goldstein (1959) developed tests containing virtually the same content but violating one of the following four accepted principles of test construction: (1) inclusion versus exclusion of irrelevant cues or specific determiners; (2) question lead versus incomplete statement lead; (3) equal-length alternatives versus extra-long alternatives; and (4) consistency versus inconsistency of grammar between lead and alternatives. In analyzing their results they found that tests constructed with faults yielded higher mean scores, which also implies that item difficulty was lowered. Validities of the tests used were not influenced by the presence of testwiseness cues. Kuder-Richardson reliability estimates, likewise, were unaffected. As the authors conclude, "no significant differential effect on reliability or validity could be attributed to violation of any of the four (test construction) principles." (p. 177)

Logic would suggest that if students are able to raise their test scores by applying test-taking skills, item-difficulties should go down and test validity must be seriously questioned. Certainly other approaches to examining the effects of testwiseness on test validity must be explored before the Dunn and Goldstein finding can be completely accepted.

Problems in Past Research on Testwiseness

Although testwiseness has been a topic of interest to educators for a number of years, relatively few studies have been devoted to empirically investigating it. The studies that have been conducted have been sporadic, often not building upon the results of previous studies.

It is difficult to derive a clear set of conclusions from past research that could serve as guidelines to practitioners. It would be useful to identify some of the problems that have plagued research in this area so that future researchers could be mindful of them in planning new studies.

1. There has been a lack of agreement as to the definition of testwiseness. Most experts believe that testwiseness represents a source of error variance that detracts from the validity of the instrument being used. However, other experts still contend that testwiseness is a general trait; akin to intelligence, and should not be separated from the skill or ability being assessed.
2. Because of the lack of a clear definition, many researchers have failed to give a complete description as to just what aspect of testwiseness they were investigating.

3. Experimental treatments have often not been fully explained. The researcher may refer to "an instructional unit in test-taking skills" without spelling out the objectives of such instruction. Also, because of the great variation in the nature and types of treatments, it is difficult to generalize across studies.

4. A great variety of criterion measures have been used to assess the effects of experimental treatments. Some investigators have constructed their own tests of testwiseness consisting of faulty items; others have relied on teacher-developed content tests; while still others have employed standardized aptitude and achievement tests. It is difficult to generalize across studies when posttests vary this much. Will the statistically significant gains demonstrated on a project-developed measure of testwiseness have any educational significance on standardized tests used in the routine school program?

Summary of the Literature on Testwiseness

After reviewing the studies that have been conducted in the area of testwiseness, it would be useful to summarize our present state of knowledge. It is difficult to derive a set of conclusions because of some of the problems cited in the previous section-wide variation in treatments and criterion measures. Nonetheless, the following conclusions are offered on a tentative basis:

1. Testwiseness does exist and can be assessed.

2. Testwiseness appears to be a construct comprised of specific skills rather than a global or general trait.

3. There is no relationship between testwiseness and sex.

4. There is no evidence that testwiseness is related to race or socioeconomic status or that students of any particular racial

group or socioeconomic level would profit more from instruction in testwiseness. A number of experts still contend that lower ability students and/or culturally different students would benefit more from such instruction, but the evidence to support their contentions does not exist.

5. Testwiseness can be effectively taught to students of all ages from preschool through adults. Although the results have not always been statistically significant, the gains nearly always favor the group receiving instruction.

6. A wide variety of methods have been successful in teaching testwiseness. Effective methods have included (1) verbally telling examinees about particular testwiseness cues; (2) providing practice in applying testwiseness cues to faulty items; (3) letting examinees read about selected aspects of testwiseness; and (4) completing self-instructional programmed exercises.

7. The long-term effects of instruction in testwiseness are doubtful, at least with the methods of instruction that have been used. Periodic instruction and distributed review would be advisable.

Table 1 contains a summary of all the studies reported in this review of the literature.

Directions for Future Research on Testwiseness

Even though past research has demonstrated that testwiseness does exist, that it can be measured, and that it can be effectively taught to relatively naive subjects, a number of unanswered questions still exist. Answers to these questions would provide more specific directions to practitioners.

1. What are the specific skills or components of testwiseness? Are some skills more important than others? Is Millman's theoretical framework adequate?

2. What are the correlates of testwiseness? Is testwiseness related to age? grade level? reading ability? socioeconomic level?

3. What effects does testwiseness have on item difficulty? item discrimination? test validity? test reliability?

Table 1 Summary of Testwiseness Literature

Study	Grade	Treatment	Tests	Results
Buchanan (1968)	Disadvantaged and non-dis. at 3 grade levels		Project developed	no sig. diff. on matched questions, more guessing by dis. on non-matched
Callenbach (1973	2	Project developed booklet	Stanford Reading Achievement Test	sig. diff. immediately post sig. diff. after 4 mos.
Diamond & Evans (1972)	6		Lorge Thorndike Iowa Test of Basic Skills Project Developed Test of test-wiseness	Ss were test-wise testwiseness not general trait testwiseness not related to general cognitive ability
Dunn & Goldstein (1959)	Army trainees		Four peoject developed tests with varied cues implanted	Intelligence & testwiseness not related
Ferrell (1972)	High school		Teacher made test; Project developed test of testwise-ness Standardized test scores	Sig. relation bet. testwiseness and teacher made test and stan. tests No. consis. reln. bet. tw and sex, or GPA
Gaines & Jongsma (1973)	5	Project developed booklet	Comprehensive Test of basic skills	Sig. increase after training

Table 1 (Con't.)

Study	Grade	Treatment	Tests	Results
Gibb (1964)	College	Training in recog. cues	Project developed test of testwiseness	Testwiseness composed in part of secondary cues Testwiseness can be taught
Kreit (1967)	3	Repeated testing	Variety of group IQ tests	Sig. improvement over 1st test; not over subsequent no sig. rel. bet. TW and intelligence
Langer, Wark & Johnson (1973)	College	Project developed material in lecture, booklet, and reading form	Project developed	Tw exists; No sig. diff. bet. methods of teaching; Sig. diff. bet. exp. and control
Moore (1971)	College	Project developed booklet	Project developed test of analogies	Sig. incr. after treatment
Lewis & Beggs (1971)	6	Repeated testing	Lorge Thorndike IQ	Sig. diff. varied bet. verbal and non-verbal
Oakland (1972)	Headstart	Project developed material	Metropolitan Readings Tests	Sig. diff. imm. post; No. sig. diff. after 4 mos.
Sause & Greico (1973)	6	Project developed material	Otis Lennon IQ	Exp. sig. higher
Slakter (1969)	8	Project developed material	Project developed	Sig. inc. if guess

Table 1 (Con't.)

Study	Grade	Treatment	Tests	Results
Slakter, Koehler, & Hampton (1970a)	5-11	Project developed material	Project developed	Sex not sig; Grade level diff. on specific cues Inc. over grade levels
Slakter, Koehler, & Hampton (1970b)	High School	Project developed programmed text	Project developed	Ss can be taught test-wiseness by programmed text
Tinney (1968)	5	Project developed material	New Developmental Reading Test	No. sign. diff.
Wahlstrom & Boersma (1968)	9	Project developed material	Project developed social studies test	Sig. diff. varied
Woodley (1973	C.L.U. insurance candidates	Project developed material	Project developed testwiseness test C.L.U. exam	Sig. diff. on test of testwiseness No. sig. diff. on C.L.U. exam

4. Do all students need training in testwiseness? Which students are likely to benefit most? When should such instruction be given?

5. What is the best method for teaching testwiseness? Should different methods be used for different age groups?

6. Why haven't the effects of instruction in testwiseness been lasting? Is it due to the instructional methods used? What can be done to produce long-term results?

7. What types of tests are most sensitive to increased testwiseness? Are the effects of instruction in testwiseness generalizable across different types of tests?

REPORT OF THE PROJECT STUDIES

As part of this project, two studies were conducted to determine the effects of a unit of instruction in test-taking skills on the reading achievement scores of fifth grade students. The same unit of instruction was administered to students at two schools in the Metropolitan New Orleans area. The schools were specifically selected to represent populations of different socioeconomic levels and racial composition. One study was conducted at a suburban elementary school and another at an inner-city elementary school. The suburban school was predominantly white (84%) with students from middle class families. The inner-city school was 56% black with students coming primarily from lower middle class homes. These two schools were chosen in an effort to find out if the results of instruction in testwiseness might differ with varied school populations and school settings.

The studies were limited to the fifth grade for two reasons. First, similar studies had been done at the kindergarten and second grade levels. A study at a higher grade level would add to the existing literature. Second, recent reading achievement test scores were available for fifth graders. These existing scores would be helpful in stratifying the sample and in facilitating additional types of data analysis.

The studies sought to answer the following questions:

1. What effects would instruction in test-taking skills have upon student performance on standardized reading achievement tests?

2. Would instruction in test-taking skills be more beneficial to students of high reading ability or students of low reading ability?

3. Would instruction in test-taking skills be more effective with lower-class students or with middle-class white students?

4. Would instruction in test-taking skills be equally effective on different standardized reading achievement tests?

Each study is reported separately in the following sections. The overall results are then discussed in the final section of this report.

Study I: Suburban Elementary

Sample

The particular suburban school used in this study had four fifth grade classes. All four classes were included in the study. Students were divided into two groups, high achievers and low achievers, on the basis of their Total Reading scores on the California Test of Basic Skills, Level 1, Form S, which had been administered in the spring of the previous school year (4.7) at the time the test was administered. High achievers were defined as those students who scored at or above grade level (4.7) at the time the test was administered. Low achievers scored below grade level. Random assignment of students to experimental and control groups was then made from each of the achievement level groups. Due to absences at the time of administration of treatment and/or at final testing time, the groups were not identical in number. The experimental group consisted of 40 students and the control group 51.

Treatment Procedures

The experimental treatment consisted of studying an instructional unit, "Test-Taking Tips," which had been developed for the project. A complete copy of the unit is included in the Appendix of this report. The unit is composed of the following sections:

1. Motivation
2. Following directions
3. Understanding what is read
4. Guessing
5. Using answer sheets correctly
6. Using time correctly
7. Test-taking conduct

This unit contains humorous illustrations and written activities designed to maintain student interest and to increase motivation and understanding. The unit is to be read by both teachers and students. Group discussion of the content is encouraged to promote better understanding for poor readers. It takes approximately one hour to one hour and 15 minutes to complete the unit.

Students in the experimental group were randomly divided into two sections and taught the unit by two of the regular classroom teachers who had volunteered to conduct the instruction. The two teachers met with the investigators on the day preceding the administration of the unit to review its contents and receive training in the use of the unit. This was done to help assure some uniformity of method in the presentation of the unit.

To control for the Hawthorne effect, the control group viewed an educational film that was unrelated to testwiseness under the supervision of the other two regular classroom teachers while the experimental group was studying the unit. After viewing the film they were encouraged to draw pictures of things seen in the film. These pictures were later collected.

Criterion Measure

On the two days following the treatment, the Reading battery of the Stanford Achievement Test, Intermediate Level I, Form A, was administered

to all fifth graders in their regular reading groups. The tests were administered by two research associates affiliated with the project. Outside test administrators were used to ensure that standardized procedures were followed. The Stanford Achievement Test was used because it was not part of the school's regular testing program and the investigators were relatively certain that none of the students had been exposed to it. The sections of the test that were used were Vocabulary, Reading Comprehension and Word Study Skills. A Total Reading score is obtained by combining the latter two subtests.

Results

The tests were machine-scored by the publisher and all data analyses were conducted using raw scores. Table 2 presents the means and standard deviations for each subtest for each of the four groups, i.e. experimental-high achievers, experimental-low achievers, control-high achievers, control-low achievers. Also found in Table 2 are the raw score differences between experimental and control groups. On each of the three subtests and Total Reading, the mean scores of the experimental group exceeded those of the Control group. An examination of the differences in raw scores reveals that high achievers receiving testwiseness instruction outscored their counterparts, who didn't receive instruction, more than experimental low achievers were able to do to their counterparts.

Table 3 presents the results of a two-way analysis of variance, with main effects of treatment and reading achievement level. As expected, there was a significant difference ($P < .001$) between high and low achievers, on each subtest and Total Reading. However, there were no significant differences between treatment groups on any of the subtests, eventhough

Table 2 SAT raw score means, standard deviations, and differences between means with breakdown by treatment and achievement levels

Group	N	SAT Vocab. $\bar{X}$	SAT Vocab. S.D.	SAT Comp. $\bar{X}$	SAT Comp. S.D.	SAT Word Study $\bar{X}$	SAT Word Study S.D.	SAT Total Rdg. $\bar{X}$	SAT Total Rdg. S.D.
Experimental									
High	16	33.56	8.16	42.44	12.61	41.25	11.08	83.69	19.86
Low	24	20.63	6.38	23.58	7.58	23.58	12.10	47.17	17.24
Total	40	25.80	9.53	31.13	13.51	30.65	14.51	61.78	25.60
Control									
High	17	30.29	5.42	38.59	9.12	41.06	8.47	79.65	11.92
Low	34	19.74	6.97	23.06	7.44	21.15	8.90	44.21	13.31
Total	51	23.25	8.17	28.24	10.86	27.78	12.85	56.02	21.15
Grand Total	91	24.37		29.51		29.04		58.55	
Difference									
High		3.27		3.85		.19		4.04	
Low		.89		.52		2.43		2.96	
Total		2.55		2.89		2.87		5.76	

Table 3 Analysis of variance of SAT scores for treatment groups by achievement levels

Test	Source of Variation	df	Mean Square	F	Significance
Vocabulary	A(Treatment)	1	69.78	1.51	n.s.
	B(Achievement level)	1	2840.97	61.64	p<.001
	A x B	1	29.41	.64	n.s.
	Within	87	46.09		
	Total	90			
Rdg. Comp.	A(Treatment)	1	68.54	.87	n.s.
	B(Achievement level)	1	6088.33	77.15	p<.001
	A x B	1	57.45	.73	n.s.
	Within	87	78.92		
	Total	90			
Word Study	A(Treatment)	1	57.61	.56	n.s.
	B(Achievement level)	1	7463.49	72.34	p<.001
	A x B	1	26.20	.25	n.s.
	Within	87	103.17		
	Total	90			
Total Rdg.	A(Treatment)	1	251.83	1.05	n.s.
	B(Achievement level)	1	27033.68	112.68	p<.001
	A x B	1	6.06	.03	n.s.
	Within	87	239.91		
	Total	90			

Table 4 Analysis of variance of SAT scores of experimental group for teacher (unit administrator) by achievement levels

Test	Source of Variation	df	Mean Square	F	Significance
Vocabulary	A(Teacher)	1	49.53	.95	n.s.
	B(Achievement level)	1	1470.86	28.17	$p < .001$
	A x B	1	8.04	.15	n.s.
	Within	36	52.22		
	Total	39			
Rdg. Comp.	A(Teacher)	1	137.57	1.39	n.s.
	B(Achievement level)	1	3096.43	31.23	$p < .001$
	A x B	1	.57	.01	n.s.
	Within	36	99.16		
	Total	39			
Word Study	A(Teacher)	1	174.30	1.26	n.s.
	B(Achievement level)	1	2680.36	19.41	$p < .001$
	A x B	1	66.04	.48	n.s.
	Within	36	138.07		
	Total	39			
Total Rdg.	A(Teacher)	1	621.57	1.85	n.s.
	B(Achievement level)	1	11538.57	34.40	$p < .001$
	A x B	1	54.32	.16	n.s.
	Within	36	335.47		
	Total	39			

the experimental group scored higher than the control group in all instances. Interaction effects between treatment and achievement level were also not significant.

Since two teachers were involved in administering the instructional unit, the investigators questioned whether differences in their manner of presentation could have influenced the effectiveness of the unit. An additional two-way analysis of variance with main effects of teacher and achievement level, was conducted to examine this issue. Table 4 contains the results of this analysis. No significant differences were found between the groups receiving instruction from the two different teachers. Teacher by achievement level interaction effects were also not significant. Once again, high achievers performed significantly better ($p. < .001$) than low achievers.

Discussion

The results seem to indicate that the students who received instruction in test-taking skills profited from the instruction. In all cases, the means of the experimental group exceeded the means of the control group. However, the differences between the two groups were not statistically significant.

Perhaps a distinction needs to be made between statistically significant differences and educationally significant differences. By examining Table 2, it can be noticed that experimental students scored approximately three raw score points higher on each of the subtests (Vocabulary, Reading Comprehension, and Word Study Skills) and nearly six points higher in Total Reading. What effect do these raw score differences have on derived scores? For students scoring at the mean or one standard deviation below,

a raw score difference of three points translates into a grade equivalent difference of four months. That is, average and slightly below average students who received instruction in test-taking skills scored approximately four months higher than similar students who did not receive instruction. For students scoring one standard deviation above the mean, this difference is closer to seven months.

Because of the inherent weaknesses and problems with grade equivalent scores, the analysis just described must be viewed cautiously. However, grade equivalent scores are the most widely used type of derived score. If grade equivalent differences of four to seven months can be produced by one hour of instruction, the results of this study appear to be educationally meaningful, if not statistically significant.

One other trend can be noted. Although the treatment by achievement level interaction was not significant, it appears that better readers profit more from instruction in test-taking skills than poor readers. Examination of the mean differences in Table 2 reveals that the differences were consistently greater for high achievers. The classifications of "high" and "low" achievers are defined by a mean-split in this study. This trend needs to be verified with a sharper division between the two achievement levels.

Study II: Inner-City Elementary

Sample

The school used in this study is part of a large urban school system. The school is located in a lower middle class neighborhood and the majority of the students are black. There were only two fifth grade classes in this school. With such a relatively small sample, no attempt was made to stratify the students on the basis of previous reading achievement scores. Previous test results revealed that, on the average, students in this school score slightly below the mean for large city populations.

All fifth grade students were randomly divided into two groups and one group was randomly assigned to receive the experimental treatment. Due to absences at the time of instruction and/or the time of testing, some students were eliminated from the study. The final sample consisted of 24 students in the experimental group and 30 students in the control group.

Treatment Procedures

The experimental treatment consisted of studying the unit "Test-Taking Tips," which has already been described and is included in the Appendix.

One of the fifth grade classroom teachers volunteered to teach the unit. The investigators met with the teacher to review the contents of the unit and to make teaching suggestions. Unlike the first study, the unit was split in half and taught on two consecutive days. The overall amount of time spent on instruction between the two studies was about equal.

While the experimental group was receiving instruction, the control group viewed educational films and filmstrips. The other fifth grade teacher conducted the activities for this group.

Criterion Measure

On the day following the treatment, all students were administered the Reading section of the Comprehensive Test of Basic Skills, Level 2, Form S, in their regular classroom. A test different from the one used in the first study was chosen as the criterion measure in order to assess the generalizability of the effects of testwiseness instruction to more than one achievement test.

Level 2 of the CTBS is intended for grades 4.5 to 6.9. The reading battery that was used consists of subtests in Vocabulary, Reading Comprehension, and Reference Skills. The first two subtests combine to yield a Total Reading score.

Results

Raw scores were used for all data analyses. Table 5 presents the means, standard deviations, and differences between the means, for each subtest and Total Reading. The experimental group scored higher, on the average, than the control group on each of the subtests as well as Total Reading.

A one-way fixed factor analysis of variance was used to test the differences between the means of the experimental and control groups for each of the subtests as well as Total Reading. These results are presented in Table 6. Although the means of the experimental group exceeded the means of the control group for each subtest, none of the differences were statistically significant ($p < .05$).

Table 5

CTBS raw score means, standard deviations, and differences between means with breakdown by treatment levels

Group	N	CTBS Vocab.		CTBS Comp.		CTBS Reference		CTBS Total Rd	
		x	S.D.	x	S.D.	x	S.D.	x	S.D.
Experimental	24	20.17	9.15	23.29	9.54	11.12	4.64	43.46	17.82
Control	30	17.73	8.73	19.50	10.15	9.67	3.53	37.23	17.91
Difference (E-C)		2.44		3.79		1.45		6.23	

Table 6

Analysis of variance of CTBS scores by treatment levels

Test	Source of Variation	df	Mean Square	F	Significance
Vocabulary	Treatment	1	78.95	.99	n.s.
	Within	52	79.52		
	Total	53			
Comprehension	Treatment	1	191.70	1.96	n.s.
	Within	52	97.82		
	Total	53			
Reference	Treatment	1	28.36	1.72	n.s.
	Within	52	16.45		
	Total	53			
Total Rdg.	Treatment	1	516.70	1.62	n.s.
	Within	52	319.20		
	Total	53			

Discussion

The results of this study are in line with the results of Study I. It appears that the students who received instruction in test-taking skills profited from their instruction. In every case the experimental students outscored the control students. However, none of the mean differences were statistically significant.

The raw score differences between the experimental and control groups range from 1.5 to about 6. These raw score differences translate to grade equivalent differences of three months to eight months. Although these differences are not statistically significant they appear to be educationally meaningful.

Discussion of Overall Results

The results of the two studies which were conducted as part of this project were not as encouraging as the investigators had hoped. It was anticipated that systematic instruction in testwiseness would produce significantly higher scores on standardized reading achievement tests. In both studies, students who received instruction in testwiseness performed better, on the average, than their classmates of comparable ability, who had not received the instruction. However, the differences between the two groups were not statistically significant. That is, it was not demonstrated that the differences could be attributed to factors other than chance.

At the risk of drawing conclusions that are not supported by the findings, the investigators strongly believe in making the distinction between "statistically significant" results and "educationally meaningful" results. Educational research is predicated on finding statistically significant results. Most educational researchers, however, will freely admit that many statistically significant results are meaningless and have little relevance to educational practice. The investigators believe that the converse is also true.

The results of the two studies conducted in this project were quite similar. When the differences between raw scores were analyzed, they were not statistically significant. Yet, when these same differences were translated to derived grade-equivalent scores, they ranged from approximately three to eight months. As stated earlier in this report, grade-equivalent scores suffer from inherent weaknesses. Despite their limitations, they

still remain the most widely used type of derived score. Teachers and principals consistently use them for placement, diagnosis, assessment of growth, program evaluation, as well as other purposes. The investigators contend that differences of three to eight months are important and meaningful to practitioners, eventhough they may not be statistically significant.

One must question why statistically significant results were not obtained and what rival hypotheses could explain the lack of differences. The two studies have limitations in sampling, design, and data analysis. The comparability of experimental and control groups is always open to question. Use of previous test scores and randomization relatively assured the investigators that the groups were of comparable ability. In the research design that was employed the effects of the instructional unit were confounded with the teachers using the unit. Because similar results were obtained with different teachers in different schools, the investigators think it is unlikely that teacher differences had a significant influence, but this rival hypothesis cannot be ruled out. Also, the investigators had no indication as to the degree of testwiseness the subjects possessed prior to the study.

What advise should be given to school administrators and classroom teachers regarding testwiseness? It would be useful to look at our current state of knowledge and derive some implications for practitioners. The following summary is based both on the review of the literature and the results of the studies conducted for this project.

Implications for Practitioners

1. Recognize that testwiseness does exist. Some students are able to perform better on tests not because they possess greater knowledge of the content, but because they can utilize various test-taking cues to improve their score.

2. Recognize that testwiseness is not a global general trait but rather a network of specific skills, some of which may be dependent upon the particular format of the test being used but others that are applicable across a wide variety of tests.

3. Don't assume that testwiseness is closely related to intelligence or academic achievement. There may be a tendency to believe that "brighter" students are more testwise than "slower" students. This hasn't been confirmed by the research evidence.

4. Prepare all students for taking tests, especially standardized tests, by providing some instruction in testwiseness shortly before the tests are administered.

5. Teach those aspects of testwiseness that appear appropriate for your age group. For younger students this might involve such features as following directions, completing answer sheets, and guessing. Older students might profit from more sophisticated skills such as using secondary cues or practicing with particular formats such as verbal analogies.

6. Use an instructional format that you believe is effective and appropriate for your age group. There is no research evidence to support one type of instruction over another. Lecture, discussion, written booklets, and programmed texts have all been used effectively. Providing concrete examples to illustrate test-taking cues would be advisable.

7. Assume that some instruction in testwiseness is better than no instruction. Such instruction is likely to be beneficial to many of the students, but statistically significant differences may not be obtained.

8. Repeat instruction in testwiseness periodically, perhaps prior to subsequent standardized test administrations. Research evidence regarding the long-term effects of testwiseness instruction is lacking, therefore, periodic review is recommended.

REFERENCES

Bauer, D. H. The effect of test instructions, test anxiety, defensiveness, and confidence in judgment on guessing behavior in multiple choice question situations. Psychology in the Schools, 1971, 8, 208-215.

Buchanan, R. G. The effect of cultural deprivation on test-taking approach as indicated by response style to multiple-choice questions. Unpublished doctoral dissertation, New York University, 1968.

Callenbach, C. A. The effects of instruction and practice in nonsubstantive test-taking techniques upon the standardized reading test scores of selected second grade students. Unpublished doctoral dissertation, Pennsylvania State University, 1971.

Chasman, D. Reading tests called wasteful, damaging. Los Angeles Times, June 24, 1972.

Diamond, J. and Evans, W. An investigation of the cognitive correlates of test-wiseness. Journal of Educational Measurement, 1972, 9 145-150.

Dunn, T. F., and Goldstein, L. G. Test difficulty, validity, and reliability as functions of selected multiple choice item construction principles. Educational and Psychological Measurement, 1959, 19, 171-179.

Ebel, R. L. Measuring educational achievement. Englewood Cliffs, N. J.: Prentice-Hall, 1965.

Ebel, R. L. and Damrin, D. E. Tests and examinations. In C. W. Harris (Ed.) Encyclopedia of Educational Research, 1960, 1502-1517.

Farr, R. Reading: what can be measured? Newark, Delaware: International Reading Association, 1969.

Ferrell, G. M. The relationship of scores on a measure of test-wiseness to performance on teacher made objective achievement examinations and on standardized ability and achievement tests, to grade point average, and to sex for each of five high school samples. Unpublished doctoral dissertation, University of Southern California, 1972.

Gaines, G. and Jongsma, E. The effect of training in test-taking skills on the achievement scores of fifth grade pupils. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, 1974.

Gibb, G. B. Test-wiseness as secondary cue response. Unpublished doctoral dissertation, Stanford University, 1964.

Jacobs, S. S. An experimental analysis of answer changing behavior on objective tests. Paper presented at the annual meeting of the American Educational Research Association, New York, February, 1971.

Kreit, L. H. The effects of practice on the acquisition of test-taking skills by a selected group of third grade pupils. Doctoral dissertation. Indiana University, 1967. Dissertation Abstracts International, 68, 3312.

Langer, G., Work, D., and Johnson, S. Test-wiseness in objective tests. In P. L. Nancke (Ed.) Diversity in Mature Reading: Theory and Research, Twenty-second Yearbook of the National Reading Conference. 1973, Vol. 1, 280-287.

Lewis, E. L. and Beggs, D. L. The effects of repeated testing on verbal and nonverbal ability assessment. Paper presented at annual meeting of American Personnel and Guidance Association, Atlantic City, N. J., April, 1971.

Millman, J., Bishop, C. H., and Ebel, R. An analysis of test-wiseness. Educational and Psychological Measurement. 1965, 25, 707-26.

Moore, J. C. Testwiseness and analogy test performance. Measurement and Evaluation in Guidance, 1971, 3, 198-202.

Oakland, T. The effects of test-wiseness materials on standardized test performance of preschool disadvantaged children. Journal of School Psychology, 1972, 10, 355-360.

Roberts, S. O. and Oppenheim, D. B. The effect of special instruction upon test performance of high school students. (ERIC Document Reproduction Service #ED 053 158).

Sause, E. F. and Grieco, F. Test sophistication and the inner city child. St. Johns University, 1973.

Slakter, M. J. Generality of risk taking on objective examinations. Educational and Psychological Measurement, 1969, 29, 115-128.

Slakter, M. J., Kochler, R. A. and Hampton, S. H. Grade level, sex, and selected aspects of test-wiseness. Journal of Educational Measurement, 1970, 119-122. (a)

Slakter, M. J., Koehler, R. A., and Hampton, S. H., Learning test-wiseness by programmed texts. Journal of Educational Measurement, 1970, 7, 247-254. (b)

Stanley, J. C. Reliability. In R. L. Thorndike (Ed.) Educational Measurement, Washington, D. C.: American Council on Education, 1971, 3, 356-442.

Thorndike, R. L. Educational measurement for the seventies. In R. L. Thorndike (ed.) Educational Measurement. Washington, D. C.: American Council on Education, 1971.

Tinney, R. E. The effect of training in test-taking skills on the reading test scores of fifth grade children of high and low socioeconomic levels. Doctoral Dissertation, University of Minnesota, 1968. Dissertation Abstracts International, 1968, 69911505.

Vernon, P. E. Symposium on the effects of coaching and practice in intelligence tests. British Journal of Educational Psychology, 1954, 24, 57-63.

Wahlstrom, M., and Boersma, P. J. The influence of test-wiseness upon achievement. Educational and Psychological Measurement, 1968, 28, 413-420.

Woodley, K. K. Investigation of test-wiseness among C.L.U. candidates. (ERIC Document Reproduction Service # ED 062-606)

RELATED SOURCES

Erickson, E. Test sophistication: an important consideration, Journal of Reading, November 1972, pp. 140-144.

Ford, V. A. Everything you wanted to know about testwiseness. U. S. Dept. of H. E. W., N. I. E. (Eric Document Reproduction Service # Ed 093 912)

Goolsby, T. M., Jr. & Wray, G. A. Practice test for preprimary and beginning first grade. Athens: Georgia University. Mr. 69. 8 p. (Eric Document Reproduction Service # ED 054 219)

Halberg, A. Facing the test: English. Miami: Dade County Public Schools, 1971, 21 p. (Eric Document Reproduction Service # ED 064 720)

Jacobs, P. T. Effects of coaching on the college board english composition test. Educational & Psychological Measurement, 1966, 26, pp. 55-67.

Lindquist, E. F. & Hieronymus, A. N. Manual for administrators, supervisors, and counselors, Iowa Tests of Basic Skills. Boston: Houghton-Mifflin Co., 1964.

The art of test taking. McCall's, 1975, 102, 38.

Millman, J. & Pauk, W. How to take tests. New York: McGraw-Hill, 1969.

Newfield, J. W. & Fluitt, J. L. On test taking skills: opposing opinions of research and practice. Louisiana Schools, 1975, 53, 10-12.

Appendix

TEST-TAKING TIPS

Developed at
Louisiana State University
at New Orleans

W. George Gaines
Eugene A. Jongsma

1.0 MOTIVATION

Test time again! Every school year there are many students who become afraid and nervous when the teacher announces that they will soon have to take some important tests.

These students are often not motivated to take tests. That is, they do not desire to take tests. These students should not be afraid but should try to do well on the tests they take.

Why do students have to take tests?

Tests are important to teachers and students for a number of reasons:

1. Tests help teachers by showing them what kinds of things their students know and don't know.

2. Some tests help teachers compare their students with other students across the country.

3. Tests also help teachers see how much their students have improved in school work.

4. When students finish school, they will often have to take tests for jobs and for college. Taking tests in school will give them practice which will be useful later in life.

Are you motivated to take tests?

When the teacher announces to your class that special tests will be given soon, don't panic. Remember that the tests are important to you and your teachers. Above all, don't be a quitter!

"THE QUITTER"

Prepare yourself for taking the tests by:

1. Getting a good night's sleep before you take the tests.

2. Telling yourself that you will try and do the very best that you can when taking the test.

3. Paying attention only to the teacher giving the directions and to the test. Try not to pay attention to other students sitting around you.

2.0 FOLLOWING DIRECTIONS

"THE CAKE DROPPER"

As the cake dropper can tell you, following directions is a must! Following directions is especially important when you are taking a test. When your teacher gives special instructions for the test, be sure that you pay close attention to what she is saying. When there are directions printed in the test booklet, read them very carefully.

Can you follow directions?

Let's check up on your ability to read and follow directions. The exercise below is a special set of tasks to see how good you are at following directions. You will be given a special sheet of paper to work on.

DIRECTIONS

1. Read all of the directions before doing anything else.
2. Write your name in the upper right hand corner of the paper.
3. Fold the paper in half, length-wise.
4. Write your name on the outside of the folded paper.
5. Print the letters of the alphabet beneath your name.
6. Add these numbers and print the sum beneath the alphabet: 2, 32, 6, 90, 45, 209.
7. Open the paper up and write your teacher's name beneath your name.
8. Write your date of birth beneath your teacher's name.
9. Place your folded paper on the floor.
10. Do not do anything asked for in 2 through 9. Raise both hands high and smile. Do not say anything.

Well, it looks like some of you need to work harder on following directions. These three simple rules will help you follow directions better when you take a test.

1. Read directions carefully. This is done by reading very slowly! Let it sink in.

2. Do exactly what the directions tell you to do.

3. If after answering a few questions you think something is not right, go back and read the directions again. Be sure you understand what the directions have said.

Here's another way for you to practice following directions. Below are the directions for a real test. Read the directions carefully and try to remember what they tell you.

TEST DIRECTIONS

"This test has 50 items. Each item has four possible answers. You are to read each item carefully and choose the one answer you think is best. There is only one right answer. You have 30 minutes to work on this test. Do not write in the test booklet."

Without looking back at the directions see if you can answer these questions:

1. How long do you have to work on the test?

2. How many items are on the test?

3. How many answers should you mark for each item?

4. Can you write in the test booklet?

Now turn back and check your answers. If you got all four right, then you are reading carefully and following directions. If you missed one or more, then you still need to think even harder about what you are doing.

Here are some questions without any directions. WHAT? NO DIRECTIONS? Yes, that's right. NO DIRECTIONS. This should give you an idea of what it would be like to take a test without taking time to read directions. By studying the questions, see if you can figure out what the correct answer is--a, b, c, or d.

1. <u>easy</u> a. hard, b. simple, c. fast, d. greasy.

2. <u>big</u> a. small, b. pig, c. large, d. boy.

3. <u>light</u> a. shiney, b. kite, c. dark, d. house.

<u>What did you think the right answers were</u>?

--If you said: simple (1-b), large (2-c), shiney (3-a), then you thought the directions asked you to choose a word meaning almost the <u>same</u> as the underlined word.

--If you said: hard (1-a), small (2-a), dark (3-c), then you thought the directions asked you to choose a word meaning the <u>opposite</u> as the underlined word.

--If you said: greasy (1-d), pig (2-b), kite (3-b), then you thought the directions asked you to choose a word that <u>sounded</u> the same as the underlined word.

Remember, following directions means to:

1. Listen to your teacher.
2. Read carefully <u>all</u> written directions.
3. Go back and read the directions again if something seems wrong.

3.0 UNDERSTANDING WHAT YOU READ

Most tests involve some reading. Students have to read the directions and then read the test questions. How well you read the test can make a difference in the score you make. Here are some important points to keep in mind when reading the test:

1. Read and follow all directions very carefully.

2. After reading a question, stop and think about what is being asked. If you are not sure, read the question again. If you are not sure after the second reading, go on to the next question.

3. Be on the lookout for small but important words like only, always, all, and never which mean "with no exceptions." Words like most, generally, and may indicate that some exceptions may be possible. Here are some examples:

   --Which of the following animals never builds a nest?

   a. Bird
   b. Beaver
   c. Turtle
   d. Dolphin

   --A desert may be

   a. cold.
   b. dry.
   c. hot.
   d. all of the above.

4. For some questions, you may be able to skim the story to find the answer. Instead of reading the whole story carefully, quickly glance through the story until you come to the part that answers the question. Here is an example of a question where you can skim to find the right answer:

   --How much water did Americans use each day in 1940?

   a. 30 billion gallons
   b. 300 billion gallons
   c. 92 billion gallons
   d. 475 billion gallons

We are now using more water than we ever did. In 1890, Americans needed only 30 billion gallons of water a day; in 1940, we were consuming 92 billion gallons daily. Today, we use approximately 300 billion gallons, however, by 1975, we will need 453 billion gallons to meet our daily needs. At that time, we may be forced to restrict the amount of water each person can use, if we wish to evade crisis situations. Thus, you can see that the growing scarcity of water is a serious problem. Scientific research must continue to investigate alternative sources to existing water supplies. Water purification machinery must continually be developed for the use of research scientists exploring ways to evade this impending crisis.

--How much water will be needed daily in the United States by 1975?

a. 30 billion gallons
b. 92 billion gallons
c. 175 billion gallons
d. 453 billion gallons

5. You may find some words on the test which you do not understand. Many times you can figure out what these words mean by looking at the other words around the unknown word to see how that word is used in that sentence or story. Here are some examples;

--The Evans family lives in a large residence on Main Street.
residence means:

a. office
b. school
c. park
d. house

--Mr. Brown told the class an old fable about "Jack and the Beanstalk."
fable means:

a. poem
b. story
c. song
d. review

4.0 GUESSING

"THE WILD GUESSER"

When students take tests there are some test questions which they are not able to answer. This is not surprising since the people who make tests include some very difficult questions which they really do not expect all students to be able to answer.

Should you guess the answer when you come to a question you don't know?

If you guess the answer to a question you should do so carefully.

Here's how!

1. After reading the question, read each of the possible answers.

2. Some of the answers will look wrong to you. Eliminate those answers.

3. This should leave you with two or three possible answers which may seem to be correct.

4. If you can limit your choices to two or three possible answers, but are still uncertain which one is correct, take a guess at the one you <u>think</u> is correct.

5. If you read all of the possible answers, and they all seem correct, <u>don't guess</u>.

"THE ELIMINATOR"

Are you like the "Eliminator" or the "Wild Guesser"?

Here are two examples which are very difficult. It is very likely that you don't know the answers to these questions. See if you can eliminate some of the choices and guess the right answer.

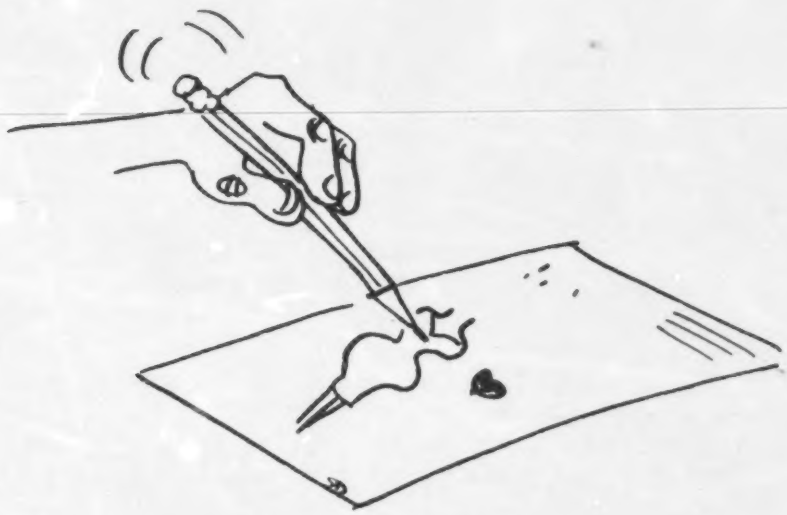
--Which word is not spelled correctly?

a. dog
b. cow
c. horsse
d. donkey

--Give means the same as

a. invest
b. go
c. contribute
d. steal

5.0 ANSWER SHEETS



"THE DOODLER"

The "Doodler" doesn't seem to know that his special answer sheet will be scored by a machine. The machine cannot tell the difference between your answers and stray marks. "Doodling" on the answer sheet may lower your score.

What is the right way to mark an answer sheet?

1. Avoid stray marks. Do not make any marks other than the answers you choose.

2. Find out which way the numbers go. Left to right? Top to bottom? Not all answer sheets are alike.

3. Keep your place on the answer sheet. Always check to make sure that your mark is put in the right place. If you are on number 23 in the test booklet, you should mark by number 23 on the answer sheet.

4. If you need to change an answer, be sure that you erase your mark cleanly.

5. Make your mark in the proper way. That is, don't use an "X" or circle, or any other mark than that called for by the directions.

6. At the end of the test go back over your answer sheet. Look to see that your erasures are clean and that no stray marks are left on the answer sheet.

See if you can tell what's wrong with these answers.

Your teacher will now tell you how to mark the answer sheet below.

6.0 TIME

"POKEY PAULA"

Time is important when taking most tests. On many tests you will be given a certain amount of time to work each section. When the time is up you must stop working and move on to the next section.

"SWIFT SARAH"

1. Your teacher will tell you how much time you will have to work on each section of a test.

    a. Looking at the classroom clock or your watch, you may figure out when that section must be completed.

    b. Before the test you may ask your teacher to write on the chalkboard the number of minutes remaining several times while you are working.

2. You have to find the right speed for yourself. Don't waste time while you are working. Move quickly from one question to another.

    a. If you work too fast, you may make careless errors.

    b. If you work too slow, you may not have time to finish.

3. While you are working, check the clock or chalkboard to see how much time you have left.

    a. If you are running out of time and have many questions left to do, you will need to work faster.

    b. If you have only one minute left but many more problems to work, you may skim over the remaining problems. Try to answer the easier ones you are sure of.

7.0 TEST-TAKING CONDUCT

A final word--try to be considerate of your classmates during the test. Be as quiet as you can. Don't make any unnecessary noise or comments. If you have a problem or need something, raise your hand and your teacher will come to see what it is. If you must speak, always do so in a low or soft voice. If you finish before time is up, go back and check your answers rather than shuffle your feet and squirm in your chair. Above all, be considerate of your classmates who are still working.

"THE YAKERS"